Reliable Entanglement Verification

Juan Miguel Arrazola,[1] Oleg Gittsovich,[1] John Matthew Donohue,[1]
Jonathan Lavoie,[1] Kevin J. Resch,[1] and Norbert Lütkenhaus[1]

[1] *Institute for Quantum Computing and Department of Physics and Astronomy,
University of Waterloo, 200 University Avenue West, N2L 3G1 Waterloo, Ontario, Canada*

(Dated: February 6, 2013)

Any experiment attempting to verify the presence of entanglement in a physical system can only generate a finite amount of data. The statement that entanglement was present in the system can thus never be issued with certainty, requiring instead a statistical analysis of the data. Because entanglement plays a central role in the performance of quantum devices, it is crucial to make statistical claims in entanglement verification experiments that are reliable and have a clear interpretation. In this work, we apply recent results by M. Christandl and R. Renner [1] to construct a reliable entanglement verification procedure based on the concept of confidence regions. The statements made do not require the specification of a prior distribution, the assumption of independent measurements nor the assumption of an independent and identically distributed (i.i.d.) source of states. Moreover, we develop numerical tools that are necessary to employ this approach in practice, rendering the procedure ready to be applied to current experiments. We demonstrate this technique by analyzing the data of a photonic experiment generating two-photon states whose entanglement is verified with the use of an accessible nonlinear witness.

PACS numbers: 03.67.-a, 03.65.Ud, 03.67.Mn

I. INTRODUCTION

Entanglement plays an essential role in various quantum information processing tasks [2–5] and experimental verification of entanglement is crucial for testing and characterizing quantum devices such as sources and channels [6]. As these devices move closer to the realm of practical technologies, our ability to perform reliable entanglement verification tests becomes increasingly important. Correspondingly, many theoretical and experimental procedures for entanglement verification have been proposed (see [2, 5] and references therein) and their improvement and development remains an active area of research [7, 8].

Any entanglement verification procedure can be thought of as a series of measurements on a physical system followed by an analysis of the outcomes. The data obtained from these measurements is necessarily finite and therefore the claim that entanglement was present can never be issued with certainty. More precisely, there will always be a non-zero probability that the data was produced from a separable state, regardless of what the data may be. We are thus forced to provide statistical statements that quantify our confidence that entanglement was indeed present. Naturally, the procedure that leads to these statements should have a clear interpretation, should not rely on unwarranted assumptions about state preparation and be readily implementable in practice [9].

The most widely used approach consists of computing the standard deviation of measured quantities and using these as error bars to specify the uncertainty of the reported values [5]. However, there are several conceptual issues with this approach [10, 11], including the fact that it can lead to counter-intuitive results [12] and is known to be inadequate to deal with nonlinear expressions [13]. This strongly asks for better alternatives and consequently other approaches have been recently suggested (see e.g. [14]).

In this paper, we apply the work of Christandl and Renner on quantum state tomography [1] to formulate a reliable method for analyzing the data of entanglement verification experiments. As shown in Ref. [1], the method does not rely on the specification of a prior distribution of prepared states nor on the assumption that they are independent and identically distributed. Additionally, it is suitable for experiments performing arbitrary quantum measurements and the final statements have a clear and well-defined operational interpretation. The approach relies on the concept of *confidence regions*: regions of state space that contain the true state with high probability [1].

Applying this method requires the specification of a region of state space for all possible measurement outcomes, an issue that is not dealt with directly in Ref. [1]. In our work we provide a recipe to assign confidence regions to data obtained from entanglement verification experiments that rely on entanglement witnesses. This assignment requires the evaluation of a non-trivial inequality for which we specifically develop numerical techniques to efficiently calculate it, rendering the entire method ready to be applied to current experiments. We demonstrate this fact by experimentally producing a family of entangled two-photon states whose entanglement is verified by an accessible nonlinear witness (ANLW) [15].

The remainder of this paper is organized as follows. For the sake of completeness, we first briefly outline the framework introduced in Ref. [1] and summarize some of its main results. We then proceed to illustrate the data analysis procedure that we build and elucidate the numerical tools that we develop to perform the necessary calculations. Finally, we describe the experimental setup and analyze the results with our technique.

II. CONFIDENCE REGIONS

We now provide an overview of the main results of Ref. [1] and direct the reader to this work for further details. We begin by considering a collection of $n+k$ quantum systems described by a state ρ^{n+k}, each system associated with a Hilbert space $\mathcal{H}$ of dimension d. The measurement is performed only on the first n systems and is described by a general POVM consisting of a set $\{B_i\}$ of positive operators satisfying $\sum_i B_i = \mathbb{1}_{\mathcal{H}}^{\otimes n}$. In the case of independent measurements of each of the systems, each element B_i will be a tensor product of n positive operators acting on a single copy of the state. However, it must be clear that the formalism does not require this assumption: one should always think of this POVM as an arbitrary, generally collective measurement on $\mathcal{H}^{\otimes n}$. The role of the remaining k systems is purely operational: the goal of the entanglement verification procedure is to make predictions about the state of these remaining systems. More precisely, we want to know if these systems belong to regions of state space that contain only entangled states. Note that n is the *number of runs* of the experiment, producing n systems which are then measured and the outcomes analyzed to build the predictions.

Consider an experiment in which the predictions are made only for a subset of k' subsystems, $k' < k$. It was noted in Ref. [1] that in the limit of $k \to \infty$, the reduced state of the $n+k'$ subsystems $\rho^{n+k'} = \mathrm{Tr}_{k-k'}(\rho^{n+k})$ can always be described by a permutationally-invariant state of the form $\int P(\sigma)\sigma^{\otimes(n+k')}d\sigma$ [16]. This corresponds to the usual independent and identically distributed (i.i.d) case in which many copies of a true state σ are prepared according to some initial probability distribution $P(\sigma)$. Thus, in the scenario of an experiment that can in principle be repeated an arbitrary number of times ($k \to \infty$) and predictions are made on a sample of k' states, the above result in fact provides a justification of the i.i.d. assumption that is common in the literature. For convenience, we will adopt this point of view but remind the reader that the i.i.d. assumption is not necessary for the validity of the upcoming results [1].

The data analysis procedure we will employ is a mapping that assigns a particular region of state space to every possible measurement outcome. Crucially, this mapping must be specified before the experiment is carried out. The regions are deemed *confidence regions* if they contain the true state with high, user-specified probability. More precisely, for all i, denote by $R(B_i)$ the region assigned to outcome B_i. This region will be a subset of the space of density matrices $\mathcal{D}(\mathcal{H})$ associated to $\mathcal{H}$. Then any prescribed region $R(B_i)$ is deemed a confidence region with confidence level $1-\epsilon$ if it satisfies the property

$$\mathrm{Prob}_{B_i}\left[\sigma \in R(B_i)\right] \geq 1-\epsilon, \quad \forall\sigma, \tag{1}$$

where $\mathrm{Prob}_{B_i}\left[\sigma \in R(B_i)\right]$ is the expected probability of success with respect to the distribution $\mathrm{Tr}\left(\sigma^{\otimes n}B_i\right)$ of the measurement outcomes B_i. In this picture, statistical statements take the following form: "We have applied a procedure that, with probability at least $1-\epsilon$, assigns a region containing the prepared state σ". It is important to emphasize that this probability refers to the success of the procedure before any measurements are carried out: in the end, the original input state σ is either definitely contained in the assigned region or not. The quantity $1-\epsilon$ should thus be interpreted as the confidence level of the statement that the state is contained in the assigned region. This statement is valid for all possible states and outcomes and does not depend on extra assumptions about state preparation nor on the prior distribution $P(\sigma)$. This fact makes the procedure reliable and robust even in the cryptographic scenario in which σ might have been chosen maliciously [1].

A main result of Ref. [1] was to provide a criteria to determine whether a given mapping from outcomes to regions succeeds in constructing confidence regions. This result is summarized as follows. Firstly, for each measurement outcome define the function

$$\mu_i(\sigma) = \frac{1}{\mathcal{N}}\mathrm{Tr}\left(\sigma^{\otimes n}B_i\right) = \frac{1}{\mathcal{N}}\mathcal{L}_i(\sigma), \tag{2}$$

where

$$\mathcal{N} = \int_{\mathcal{D}(\mathcal{H})} \mathcal{L}_i(\sigma)d\sigma$$

is a normalization constant. The function $\mathrm{Tr}\left(\sigma^{\otimes n}B_i\right)$ is usually referred to as the *likelihood function*, so that $\mu_i(\sigma)$ is simply its normalized version. Furthermore, let $\{\Gamma_i\}$ be a collection of subsets of $\mathcal{D}(\mathcal{H})$, where the number of these regions is equal to the number of POVM elements $\{B_i\}$. For each set Γ_i define the enlarged set

$$\Gamma_i^\delta = \{\sigma : \exists\sigma' \in \Gamma_i \text{ such that } F(\sigma,\sigma') \geq \sqrt{1-\delta^2}\}, \tag{3}$$

where $F(\sigma,\sigma') = \mathrm{Tr}\left(\sqrt{\sqrt{\sigma}\sigma'\sqrt{\sigma}}\right)$ is the fidelity and

$$\delta^2 = \frac{2}{n}\left[\ln\frac{2}{\epsilon} + (d^2-1)\ln n\right]. \tag{4}$$

If for all possible outcomes B_i it holds that

$$\int_{\Gamma_i} \mu_i(\sigma)d\sigma \geq 1 - \frac{\epsilon}{c_{n,d}} \tag{5}$$

with

$$c_{n,d} = 2n^{(d^2-1)/2}, \tag{6}$$

then the assigned regions Γ_i^δ are confidence regions with confidence level $1-\epsilon$ (Corollary 1, [1]). In equation (5),

$d\sigma$ is the Hilbert-Schmidt measure: the flat measure on the set of density matrices of dimension d induced from the Haar measure on the set of pure states of dimension $d \times d$ [17]. It must be noted that the polynomial factor $2n^{(d^2-1)/2}$ [18] is an improvement on the term appearing in Ref. [1].

The above condition (5) can be more conveniently cast by referring directly to the quantity $1 - \int_{\Gamma_i} \mu_i(\sigma) d\sigma$ and making a direct comparison with the term $\epsilon / c_{n,d}$. This can be achieved by instead integrating over the complement regions $\overline{\Gamma_i} = \{\sigma : \sigma \notin \Gamma_i\}$. Therefore we define

$$\begin{aligned} \epsilon_2(B_i, \Gamma_i) &:= \int_{\overline{\Gamma_i}} \mu_i(\sigma) d\sigma \\ &= \frac{\int_{\overline{\Gamma_i}} \mathcal{L}_i(\sigma) d\sigma}{\int_{\mathcal{D}(\mathcal{H})} \mathcal{L}_i(\sigma) d\sigma}. \end{aligned} \tag{7}$$

For convenience, we will drop the explicit dependence on B_i and Γ_i from $\epsilon_2(B_i, \Gamma_i)$ whenever it is not necessary, while keeping in mind that its value will depend on the measurement outcome and the region assigned to it. Condition (5) can then be more conveniently cast as

$$\epsilon_2 \cdot c_{n,d} \leq \epsilon. \tag{8}$$

In summary, the assigned regions $\{\Gamma_i\}$ determine whether criteria (8) is satisfied for some fixed value of ϵ and whenever it is, the enlarged regions Γ_i^δ constitute confidence regions. It is these latter regions that we assign to each individual outcome in our data analysis procedure.

It is very important to note the role played by the polynomial factor $c_{n,d}$ and the enlarging parameter δ. Because the dimension of the Hilbert space d is fixed for a given experiment and typically large, the factor $c_{n,d}$ will be a high-order polynomial in the number of runs n. Satisfying condition (8) will require ϵ_2 to be much smaller than the value of ϵ that quantifies the confidence of the procedure. This can be problematic for small n but will play only a minor role for larger values because ϵ_2 decreases exponentially in n whenever the maximum of the function $\mu_i(\sigma)$ is contained in the region Γ_i [1].

On the other hand, the size of the complement region $\overline{\Gamma_i}$ increases as δ grows larger, implying that large values of δ result in larger values of ϵ_2. In particular, whenever $\delta \geq 1$ (which can occur for sufficiently low n) it will hold that the region $\overline{\Gamma_i}$ will be equal to the entire state space $\mathcal{D}(\mathcal{H})$ and consequently $\epsilon_2 = 1$. Thus, for a fixed confidence level, the value of n for which $\delta = 1$ sets a lower limit on the number of runs of the experiment that are required to verify the presence of entanglement. This is illustrated in Fig. 1. These features indicate that in this framework, it is usually necessary to accumulate large amounts of data in order to reliably report the presence of entanglement.

We have in hand a method to verify whether a set of prescribed regions are in fact confidence regions. The question then remains of how to choose these regions in the first place, an issue that is not addressed in Ref. [1].

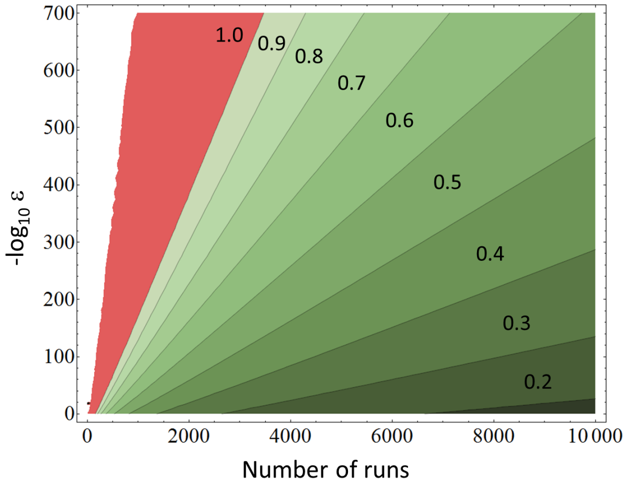


FIG. 1: (Color online) Contour plot of δ as a function of the confidence and number of runs n. The red region to the farmost left represents the case when $\delta > 1$, illustrating a lower bound on the number of runs that must be performed to achieve a certain value of ϵ, quantified by the quantity $-\log_{10} \epsilon$. In practice, even larger values of n will be required to meet a desired confidence.

Although the results of Christandl and Renner were originally targeted at quantum state tomography, we will instead apply these results in the context of entanglement verification. We now describe a procedure for entanglement verification that fully specifies how to assign confidence regions in terms of entanglement witnesses.

III. ENTANGLEMENT VERIFICATION PROCEDURE

The goal of an entanglement verification experiment is to determine whether a prepared state is entangled or not with the highest possible certainty. In the language of confidence regions this translates to the task of deciding with the highest level of confidence possible whether the prepared state lies in a region consisting only of entangled states. Reconstructing the state of a general quantum system is experimentally demanding, as the number of required measurement settings will in general increase exponentially with the number of qubits [5]. Moreover, even if a given state is completely specified, deciding conclusively whether it is entangled is computationally demanding and it is in fact an NP-hard problem in terms of the dimension of the system [19]. A way to circumvent these issues is to focus on entanglement witnesses, the use of which has become an increasingly popular tool both in theory and experiments [20–24], thus playing a central role in the field of entanglement verification.

A linear entanglement witness W is an observable sat-

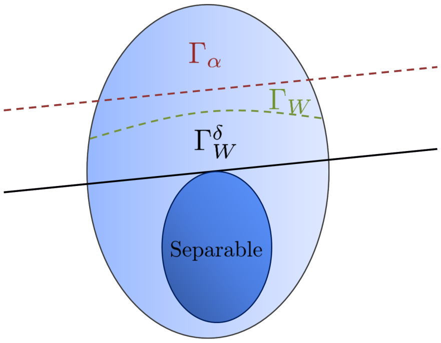


FIG. 2: (Colour online) The region Γ_W^δ is fixed as the set of states detected by a linear entanglement witness W. This region can be seen as the set of states *above* the black line. Fixing Γ_W^δ implicitly defines a region Γ_W that determines if criteria (8) is satisfied. This region is located above the dashed green line labelled Γ_W. The required numerical efforts are greatly simplified by realizing that the set of states Γ_α above the dashed red line constitute a subset of Γ_W as in Observation 1.

isfying

$$w(\sigma_s) := \mathrm{Tr}\,(\sigma_s W) \geq 0 \;\text{ for all } \sigma_s \text{ separable,}$$
$$w(\sigma_e) < 0 \;\text{ for at least one entangled state } \sigma_e.$$

Therefore, recording a negative expectation value is a conclusive indicator that the state must have been entangled. We refer to this as the state being *detected* by the witness. Calculating the expectation value of a witness operator can be performed efficiently for any state of arbitrary dimension. Moreover, experimentally determining the expectation value of a witness generally requires considerably fewer measurement settings than a full reconstruction of the state, making them very attractive in practical scenarios.

One can also consider nonlinear entanglement witnesses [25, 26] which can be viewed as powerful extensions of linear witnesses in the sense that they will always detect more states than their linear counterparts. Nonlinear witnesses are described by their values $w(\sigma)$ which are nonlinear in the expectation value of the measured observables. They also satisfy the property that their value is negative only for entangled states. Moreover, accessible nonlinear witnesses were recently developed in [15], demonstrating that their expectation value can be evaluated from the same data as the original linear witness. Such nonlinear witnesses have also been recently applied in experiments [27].

The starting point of our procedure is the specification of an entanglement witness W and a POVM $\{B_i\}$ whose possible outcomes are sufficient to determine the expectation value of W. In our description $w(\sigma)$ refers to the value of a linear or nonlinear witness. Recall that in order to verify entanglement whenever it is present, we need to assign confidence regions that contain only entangled states. For this purpose, we define

$$\Gamma_W^\delta := \{\sigma : w(\sigma) < 0\} \quad \forall i \tag{9}$$

as the set of detected states. From the definition of an entanglement witness, Γ_W^δ contains only entangled states. Our goal will be to report Γ_W^δ as the confidence region whenever possible. Going back to definition (3), notice that the set Γ_i^δ is defined for a fixed Γ_i. But in our picture, we are interested in always reporting regions that contain only entangled states. Therefore, we alternatively choose to fix the reported region Γ_W^δ and construct the smaller regions implicitly. From (3), it can be directly seen that if Γ_W^δ is fixed, its corresponding subregion Γ_W is defined by

$$\Gamma_W := \{\sigma : \max_{\sigma' \in \overline{\Gamma_W^\delta}} F(\sigma, \sigma') < \sqrt{1-\delta^2}\}. \tag{10}$$

We are now ready to specify the mapping from outcomes to regions that constitutes the data analysis procedure for reliable entanglement verification.

Data analysis procedure. To construct confidence regions with confidence level $1-\epsilon$ in an entanglement verification experiment, apply the following rule to assign a region to each outcome B_i:

1. Fix ϵ.
2. For each possible measurement outcome B_i, compute $\epsilon_2(B_i, \Gamma_W) = \int_{\overline{\Gamma_W}} \mu_i(\sigma) d\sigma$.
3. If condition (8) holds, i.e. if $\epsilon_2 \cdot c_{n,d} \leq \epsilon$, assign the set of detected states Γ_W^δ. Otherwise, assign the entire state space $\mathcal{D}(\mathcal{H})$.

Therefore, we assign only two possible regions: the set of detected states Γ_W or the entire state space $\mathcal{D}(\mathcal{H})$. Note that the entire state space is trivially a confidence region for any given confidence level, so that our assignment indeed produces confidence regions. However, assigning the entire state space must be interpreted as the statement that for the given confidence level, it is not possible to certify that the set of detected states contains the true state.

Even though the procedure is now completely specified, we are still faced with the difficulty of calculating ϵ_2. As a first step, we note that it is preferable to find a simpler way to characterize the set Γ_W. One way to do this is to find a subset of Γ_W that can be more easily described. We now show that such a subset can always be found in terms of a bound on the expectation value of a linear entanglement witness.

Observation 1. *Let W be an entanglement witness and let the number $\alpha > 0$ satisfy $\alpha > 2||W||_\infty \delta$. Then the set $\Gamma_\alpha = \{\sigma : Tr(\sigma W) < -\alpha\}$ is a subset of Γ_W.*

Proof: In order to prove the claim we only need to show that $F^2(\sigma, \sigma') < 1 - \delta^2$ whenever $\mathrm{Tr}\,(\sigma W) < -\alpha$

and $\mathrm{Tr}\,(\sigma' W) > 0$. We begin by considering the following general inequality:

$$\begin{aligned}|\mathrm{Tr}\,[(\sigma'-\sigma)W]\,| &= |\langle W, \sigma'-\sigma\rangle| \\ &\leq ||W||_\infty ||\sigma'-\sigma||_{\mathrm{tr}} \\ &\leq 2||W||_\infty \sqrt{1-F^2(\sigma,\sigma')} \end{aligned} \tag{11}$$

where we have used *Hölder's inequality*

$$|\langle\sigma, W\rangle| \leq ||\sigma||_{\mathrm{tr}} ||W||_\infty \tag{12}$$

and the *Fuchs-van de Graaf inequality* [28]

$$||\sigma'-\sigma||_{\mathrm{tr}} \leq 2\sqrt{1-F^2(\sigma,\sigma')}. \tag{13}$$

Now let $\mathrm{Tr}\,(\sigma W) = -\alpha$ and $\mathrm{Tr}\,(\sigma' W) = \beta$ for some $\alpha, \beta > 0$. Inserting into (11) and rearranging we get

$$F^2(\sigma,\sigma') \leq 1 - \left(\frac{\beta+\alpha}{2||W||_\infty}\right)^2.$$

We want to find a condition on α such that $F^2(\sigma,\sigma') < 1-\delta^2$ for any β. This will occur whenever

$$\begin{aligned} 1 - \left(\frac{\beta+\alpha}{2||W||_\infty}\right)^2 &< 1-\delta^2 \\ \Rightarrow \alpha &> 2||W||_\infty \delta - \beta. \end{aligned}$$

Since this inequality must hold for all β, we can restrict ourselves to the worst case scenario of $\beta = 0$ to obtain

$$\alpha > 2||W||_\infty \delta \tag{14}$$

as desired. ■

This result is illustrated in Fig. 2. Unfortunately, obtaining a similar and useful result for nonlinear witnesses is difficult: the value of the nonlinear witness may differ greatly for two states even if their fidelity is high.

Note that because $\Gamma_\alpha \subseteq \Gamma_W$, it holds that

$$\int_{\overline{\Gamma_\alpha}} \mathcal{L}_i(\sigma) d\sigma \geq \int_{\overline{\Gamma_W}} \mathcal{L}_i(\sigma) d\sigma \quad \forall i, \tag{15}$$

since $\mathcal{L}_i(\sigma) \geq 0$. Therefore if condition (8) is satisfied when integrating over $\overline{\Gamma_\alpha}$, it will always be satisfied for the integral over $\overline{\Gamma_W}$.

Typically, it is possible to assign the set of detected states as a confidence region for very high confidence levels i.e. with $\epsilon \ll 1$. Therefore, from now on we will quantify the confidence level of the procedure by the more appropriate quantity

$$C = -\log_{10} \epsilon, \tag{16}$$

which we refer to as the *confidence* of the entanglement verification procedure. We further define this quantity to be zero whenever the assigned region is the entire state space $\mathcal{D}(\mathcal{H})$. Thus, higher values of the confidence result in higher certainty that the state is contained in the set of detected states.

From the description of the data analysis procedure, it should be clear that the crucial step is the computation of ϵ_2: a highly non-trivial task that requires the normalization of the likelihood function as well as its integral over the implicitly defined set Γ_W. In the following section, we construct and illustrate a series of tools developed to numerically evaluate an upper bound on ϵ_2, ensuring a method to verify condition (8).

IV. NUMERICAL TOOLS

There are several difficulties in calculating ϵ_2. An analytical approach is essentially intractable owing primarily to the high dimensionality of parameter space and the non-trivial geometry of the space of positive semi-definite operators [29]. Moreover, the region of integration $\overline{\Gamma_W}$ is not known in closed form but can only be cast as a black-box i.e. we can only ask whether a state lies in this region or not. Finally, we require any approximation of ϵ_2 to provide an upper bound on its value in order to ensure that the inequality $\epsilon_2 \leq c_{n,d}\epsilon$ is always satisfied.

Fortunately, high-dimensional integration over black-box constraints can be handled with the use of Monte Carlo techniques. Most of these techniques are well summarized in [30]. In the Monte Carlo approach, the mean value of the integrand is approximated by the average value of samples randomly drawn from the region of integration, which in conjunction with knowledge of the hyper-volume of the integration region can be used to calculate the value of the integral. Importantly, the number of samples can be chosen independently of the underlying dimension and any constraint can be straightforwardly incorporated by checking whether a sample point lies within the constraint region.

More specifically, the simplest version of a Monte Carlo technique to approximate a general integral of the form $\int_R f(\sigma)d\sigma$ involves a random sequence of N density operators $\{\sigma_1, \sigma_2, \ldots, \sigma_N\}$ uniformly sampled inside R according to the measure $d\sigma$. By definition, the average $\langle f\rangle_R$ of a function over a region R satisfies

$$\int_R f(\sigma)d\sigma = \langle f\rangle_R \cdot V_R \tag{17}$$

where $V_R = \int_R d\sigma$ is the hyper-volume of the integration region. The goal in Monte Carlo integration is to approximate the average of the function from the random sample. Namely, we approximate the value of the integral as

$$\int_R f(\sigma)d\sigma \approx \left[\frac{1}{N}\sum_{j=1}^{N} f(\sigma_j)\right] \cdot V_R, \tag{18}$$

while keeping in mind that all sampled states lie in the integration region. Convergence to the true value of the integral is guaranteed as $N \to \infty$ due to the law of large numbers [30]. A main drawback of this approach is that convergence can be extremely slow for highly-peaked functions such as $\mathcal{L}_i(\sigma)$, since only very rarely will a state be drawn from the region surrounding the maximum of the function. This is particularly troublesome for our purposes because an error in the calculation of ϵ_2 can lead to wrong conclusions about the confidence of the procedure. For this reason, we now introduce an approach that can be easily and efficiently implemented and provides an upper bound on ϵ_2.

We first note that such a bound can be achieved by introducing a lower bound on the normalization constant $\mathcal{N}$. Since the likelihood function is strictly positive, this can always be achieved by integrating over a subset R of $\mathcal{D}(\mathcal{H})$, i.e.

$$\epsilon_2 \leq \frac{\int_{\overline{\Gamma_W}} \mathcal{L}_i(\sigma) d\sigma}{\int_R \mathcal{L}_i(\sigma) d\sigma}. \tag{19}$$

We can use this fact to our advantage by restricting R to be a region around the maximum of $\mathcal{L}_i(\sigma)$. Note that this maximum is unique and is in general achieved for a convex set of states [14]. Ideally, this region should be chosen to satisfy $\int_R \mathcal{L}_i(\sigma)d\sigma \approx \int_{\mathcal{D}(\mathcal{H})} \mathcal{L}_i(\sigma)d\sigma$ in order to provide a tight bound, but this is not necessary as the bound is guaranteed to hold for any R. Additionally, because the likelihood function is more flat around the maximum and R is much smaller than $\mathcal{D}(\mathcal{H})$, drawing random states within R will greatly improve the convergence of a Monte Carlo integration.

We now illustrate how this region R can be constructed from a hyper-rectangle in parameter space. Following the convention of [31], we begin by parametrizing any state $\sigma \in \mathcal{D}(\mathcal{H})$ in terms of the real-valued Bloch vector $\tau = (\tau_1, \tau_2, \ldots, \tau_{d^2-1})$ as

$$\sigma(\tau) = \frac{1}{d}\mathbb{1} + \sum_{j=1}^{d^2-1} \tau_j \hat{\lambda}_j, \tag{20}$$

where the operators $\{\hat{\lambda}_j\}$ are an orthogonal set of traceless Hermitian generators of $SU(d)$ satisfying $\mathrm{Tr}\left(\hat{\lambda}_j^{\,2}\right) = 1$. Any operator written in such a form is immediately Hermitian and of unit trace but may be non-positive for some vectors τ. Thus, it will be important to keep in mind that not all possible vectors yield valid density matrices. With this parametrization the likelihood function will be a function of the Bloch vector $\mathcal{L}_i(\sigma) = \mathcal{L}_i(\tau_1, \tau_2, \ldots, \tau_{d^2-1})$. Our goal will be to define a region around the maximum that contains only valid states for which the value of the likelihood function is sufficiently large.

Construction of integration regions. To construct a region R to be used in an approximation of the normalization of the likelihood function, perform the following:

1. Calculate the maximum value of the likelihood function $\mathcal{L}_i^{\max}$ and any vector $\tau^* = (\tau_1^*, \tau_2^*, \ldots, \tau_{d^2-1}^*)$ for which this maximum is attained.

2. Find, for all j, the lowest possible quantities $x_j^\pm > 0$ such that $\mathcal{L}_i(\tau_1^*, \tau_2^*, \ldots, \tau_j^* \pm x_j^\pm, \ldots, \tau_{d^2-1}^*) = \mathcal{L}_i^{\max}/\eta$ for some fixed number $\eta > 0$. If no such values can be found for some j, let $x_j^\pm = \infty$.

3. Find, for all j, the highest possible quantities $y_j^\pm > 0$ such that $\sigma(\tau_1^*, \tau_2^*, \ldots, \tau_j^* \pm y_j^\pm, \ldots, \tau_{d^2-1}^*)$ is still a valid density matrix.

4. Define $r_j^\pm = \min\{x_j^\pm, y_j^\pm\}$. Then the integration region R is equal to all the valid density matrices within the hyper-rectangle r defined by $r = \{\tau : \tau_j^* - r_j^- \leq \tau_j \leq \tau_j^* + r_j^+, \forall j\}$.

This construction is illustrated in Fig. 3. Note that the task of maximizing the likelihood function can be performed efficiently and is routine in the context of quantum state tomography. A good choice of η will in general depend on each individual problem, but it should be chosen large enough to include only regions that contribute significantly to the integral.

Once the hyper-rectangle has been constructed, it is straightforward to perform the Monte Carlo integration by sampling uniformly within the rectangle, while keeping only operators in that sample that are valid density matrices. Let these sampled states form the set $\{\sigma_1, \sigma_2, \ldots, \sigma_N\}$. The target integral is then given by

$$\begin{aligned}\int_R \mathcal{L}_i(\sigma)d\sigma &\approx \left[\frac{1}{N}\sum_{j=1}^{N} \mathcal{L}_i(\sigma_j)\right] \cdot V_R \\ &= \langle \mathcal{L}_i \rangle_R \cdot V_R.\end{aligned} \tag{21}$$

Because typical values of the likelihood function are extremely small, it is preferable to work with the logarithm of the function and use the identity

$$\log{(a+b)} = \log[\exp(\log a - \log b) + 1] + \log b \tag{22}$$

to add the values of $\mathcal{L}_i(\sigma_j)$ at each step of the algorithm and determine $\langle \mathcal{L}_i \rangle_R$ as in equation (21).

In order to calculate V_R, we use the fact that the Hilbert-Schmidt metric on the space of quantum states generates the Hilbert-Schmidt measure [31]. The Hilbert-Schmidt distance between two density matrices is given by

$$D_{HS}(\sigma_1, \sigma_2) = ||\sigma_1 - \sigma_2||_2 = \sqrt{\mathrm{Tr}\,[(\sigma_1 - \sigma_2)^2]}. \tag{23}$$

This correspondence between metric and measure implies that the volume of the hyper-rectangle r can be found in the usual sense as the product of the length of its sides with respect to the Hilbert-Schmidt metric. More specifically, let $\sigma_j^\pm = \sigma(\tau_1^*, \ldots, \tau_j^* \pm r_j^\pm, \ldots, \tau_{d^2-1}^*)$. Then the length Δr_j of the jth side of r is given simply by

$$\begin{aligned}\Delta r_j &= D_{HS}(\sigma_j^+, \sigma_j^-)\\ &= \|r_j^+ \hat{\lambda}_j + r_j^- \hat{\lambda}_j\|_2 = r_j^+ + r_j^-,\end{aligned} \tag{24}$$

where we have used the fact that the operators $\hat{\lambda}_j$ are normalized with respect to the Hilbert-Schmidt inner product. The hyper-volume V_r of r is then given by

$$V_r = \prod_{j=1}^{d^2-1} \Delta r_j. \tag{25}$$

This correspondence is also useful in generating a random sample, as one needs only to obtain a random number within the intervals $[\tau_j^* - r_j^-, \tau_j^* + r_j^+]$. Because not all operators in r are valid density matrices, V_r is in general larger than the hyper-volume V_R of the integration region R. However, one can estimate R from knowledge of the fraction f of the randomly drawn operators that are valid density matrices. The relationship between these quantities is

$$V_R \approx f \cdot \prod_{j=1}^{d^2-1} \Delta r_j, \tag{26}$$

which can finally be inserted in (21) to provide the numerical calculation of the target integral

$$\int_R \mathcal{L}_i(\sigma) d\sigma \approx \left[\frac{1}{N}\sum_{j=1}^{N} \mathcal{L}_i(\sigma_j)\right] \cdot f \cdot \prod_{k=1}^{d^2-1} \Delta r_k. \tag{27}$$

To calculate f, it is sufficient to verify how many of the drawn operators are valid density operators and divide this number by the total number of randomly drawn operators.

One could imagine that a similar technique could be used to calculate the integral $\int_{\overline{\Gamma_W}} \mathcal{L}_i(\sigma) d\sigma$ appearing in the definition of ϵ_2. Unfortunately, this would greatly increase the computational efforts in the construction of R since one must additionally ensure that each of the drawn samples lie in Γ_W. Additionally, in this case restricting the integration region results in an incorrect lower bound on ϵ_2. Instead, we can construct an upper bound on this integral via the maximum of the likelihood function as

$$\begin{aligned}\int_{\overline{\Gamma_W}} \mathcal{L}_i(\sigma) d\sigma &= \langle \mathcal{L}_i \rangle_{\overline{\Gamma_W}} \cdot V_{\overline{\Gamma_W}}\\ &\leq \left(\max_{\sigma \in \overline{\Gamma_W}} \mathcal{L}_i(\sigma)\right) \cdot V_{\mathcal{D}(\mathcal{H})},\end{aligned} \tag{28}$$

where $V_{\mathcal{D}(\mathcal{H})}$ is the Hilbert-Schmidt hyper-volume of the entire state space. This volume was calculated explicitly in [31] for Hilbert spaces of arbitrary dimension. We can then combine this result with our previous bound on the normalization constant to provide an overall upper bound on ϵ_2. Since this value will be typically very small and in order to make a direct comparison with the confidence, we will henceforth refer to the logarithm of ϵ_2 for which we now have the inequality

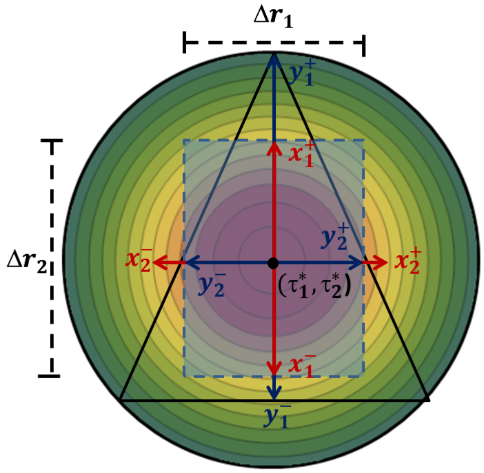


FIG. 3: (Color online) Construction of integration regions. We imagine a two-dimensional section of parameter space characterized by the variables τ_1 and τ_2. Only the region inside the triangle contains valid density matrices and contours of $\mathcal{L}_i(\sigma)$ are shown in the background. To construct the integration region we do the following: 1. Find the maximum of the function and a state for which it occurs, in this case (τ_1^*, τ_2^*). 2. From this maximum, find the displacements $x_1^\pm$ and $x_2^\pm$ such that the value of the function is decreased by a specified amount, in this case corresponding to the 6th contour line. 3. Find the displacements $y_1^\pm$ and $y_2^\pm$ that define the points where the boundary of valid states is met. 4. By choosing the minimum of these quantities in each direction, we construct a rectangle (dashed) and the integration region is the intersection of this rectangle with the space of valid density matrices.

$$\log_{10} \epsilon_2 \leq \log_{10}\left(\frac{\max_{\sigma \in \overline{\Gamma_W}} \mathcal{L}_i(\sigma)}{\langle \mathcal{L}_i \rangle_R} \frac{V_{\mathcal{D}(\mathcal{H})}}{V_R}\right). \tag{29}$$

Of course, the average of the likelihood function over $\overline{\Gamma_W}$ will generally be much smaller than the maximum over this region, making the bound very loose. However, in practice this is not a problem because the above bound on ϵ_2 is dominated by the much larger differences between the global maximum of the function and its maximum over $\overline{\Gamma_W}$. More specifically, for experiments with a large number of runs (large n), it will typically hold that

$$\left|\log_{10}\left(\frac{\max_{\sigma \in \overline{\Gamma_W}} \mathcal{L}_i(\sigma)}{\langle \mathcal{L}_i \rangle_R}\right)\right| \gg \left|\log_{10}\left(\frac{\langle \mathcal{L}_i \rangle_{\overline{\Gamma_W}}}{\max_{\sigma \in \overline{\Gamma_W}} \mathcal{L}_i(\sigma)}\right)\right|, \tag{30}$$

so that

$$\log_{10}\left(\frac{\langle\mathcal{L}_i\rangle_{\overline{\Gamma_W}}}{\langle\mathcal{L}_i\rangle_R}\right) =$$
$$\log_{10}\left(\frac{\max_{\sigma\in\overline{\Gamma_W}}\mathcal{L}_i(\sigma)}{\langle\mathcal{L}_i\rangle_R}\frac{\langle\mathcal{L}_i\rangle_{\overline{\Gamma_W}}}{\max_{\sigma\in\overline{\Gamma_W}}\mathcal{L}_i(\sigma)}\right)$$
$$\approx \frac{\max_{\sigma\in\overline{\Gamma_W}}\mathcal{L}_i(\sigma)}{\langle\mathcal{L}_i\rangle_R} \tag{31}$$

and the value for $\log_{10}\epsilon_2$ is not altered significantly by the loose bound.

The final quantity we must be able to calculate is the maximum of the likelihood function over $\overline{\Gamma_W}$. This again is a non-trivial global optimization problem involving a black-box constraint. As in the case of integration, the particular features of this problem impede the usual techniques and strongly ask for a Monte Carlo approach. To handle the optimization in the general case, we employ an adaptation to the quantum scenario of a simulated annealing algorithm (SA) based on the Metropolis-Hastings algorithm outlined in [32].

The SA algorithm is based on a biased random walk that preferentially moves to states with higher values of the objective function while still accepting moves to lower values with a probability governed by a global "temperature" parameter. This last feature prevents the algorithm from being confined in local maxima. Unfortunately, this same feature makes the convergence slow, usually requiring many steps to reach close proximity to the maximum. For each step, one must additionally make the costly verification that the states lie in the region of integration $\overline{\Gamma_W}$, so it must be understood that run times are usually long. A detailed description of the algorithm is included in the Appendix.

One drawback of the SA algorithm is that due to its stochastic nature, independent runs of the algorithm will generally yield different values. Moreover, by construction these values cannot be larger than the global maximum. In order to address this issue, one should estimate the numerical error by performing many independent runs of the algorithm and collecting statistics of the sample values. The usual choice is to calculate the standard deviation of the values [30] and take this as the error. It is then important to ensure that condition (8) is satisfied well within this error.

Nevertheless, we are still interested in obtaining a more efficient method to solve the maximization problem. We can achieve this for the case of linear witnesses by noting that for the subset Γ_α of Γ_W, it holds that

$$\max_{\sigma\in\overline{\Gamma_\alpha}}\mathcal{L}_i(\sigma) \geq \max_{\sigma\in\overline{\Gamma_W}}\mathcal{L}_i(\sigma) \tag{32}$$

since in that case $\overline{\Gamma_W}$ is a subset of $\overline{\Gamma_\alpha}$. Therefore, we can provide a final expression for the bound on ϵ_2 as

$$\log_{10}\epsilon_2 \leq \log_{10}\left(\frac{\max_{\sigma\in\overline{\Gamma_\alpha}}\mathcal{L}_i(\sigma)}{\langle\mathcal{L}_i\rangle_R}\frac{V_{\mathcal{D}(\mathcal{H})}}{V_R}\right) \tag{33}$$

where Γ_α is defined as in Observation 1. This expression has the enormous advantage that because the constraint over Γ_α is convex and $\mathcal{L}_i(\sigma)$ is log-convex, the maximization of $\mathcal{L}_i(\sigma)$ over this region can be calculated with vastly greater efficiency using standard methods in convex optimization.

We are additionally interested in reporting the highest possible confidence level, which corresponds to the case in which the equality $\epsilon_2 \cdot c_{n,d} = \epsilon$ holds. The value of ϵ_2 depends on the region Γ_W which in turn implicitly depends on ϵ through the definition of the enlarging parameter δ, so that the above equality is in principle an equation to be solved for ϵ. Unfortunately, there is no clear method of how to solve the equation directly, primarily because of the difficulty of calculating ϵ_2 itself. Instead, to achieve the highest possible confidence level, one must iteratively adapt the chosen value of ϵ until $\epsilon_2 \cdot c_{n,d} \approx \epsilon$ while still satisfying the inequality (8).

With these tools in hand it is now possible to apply the reliable entanglement verification procedure for both linear and nonlinear witnesses. We now proceed to demonstrate the features of the method by applying the technique to data obtained from an experiment generating a family of entangled two-photon states. The entanglement of these states is verified with the use of an ANLW.

V. EXPERIMENT

To apply our entanglement verification procedure to experimental data, we aimed to produce photon pairs in the maximally entangled states $|\Phi(\phi)\rangle = \frac{1}{\sqrt{2}}\left(|HH\rangle + e^{i\phi}|VV\rangle\right)$, where $|H\rangle$ and $|V\rangle$ are defined respectively as polarization parallel and perpendicular to the optical table. A frequency doubled titanium-sapphire laser (80 MHz, 790 nm) was used to pump a pair of orthogonally oriented 1 mm β-Barium borate (BBO) crystals, as seen in Fig. 4. By pumping with diagonal polarization $|D\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)$, the pump may produce photon pairs via type-I noncollinear spontaneous parametric down-conversion (SPDC) in either the first or second crystal [33]. Bismuth borate, α-BBO, and quartz crystals were used to ensure that each path was spatially and temporally indistinguishable, and the photon pairs were filtered using bandpass filters with a centre wavelength of 790 nm and a bandwidth FWHM of 3 nm. The single photon signal was measured with avalanche photodiodes (APDs) and coincidences were recorded within a 3 ns window.

Single photons were detected at a rate of approximately 200 kHz in each arm, with a coincidence rate of approximately 35 kHz when the measurements are set to HH or VV. A quarter-wave plate was tilted to introduce an arbitrary phase shift between horizontally and vertically polarized components, allowing control over the phase ϕ. This setup constitutes part of the setup used for the experiment reported in [34]. The two-photon state was prepared for six values of ϕ, corresponding to a wave-

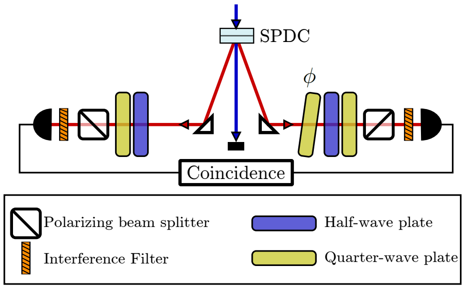
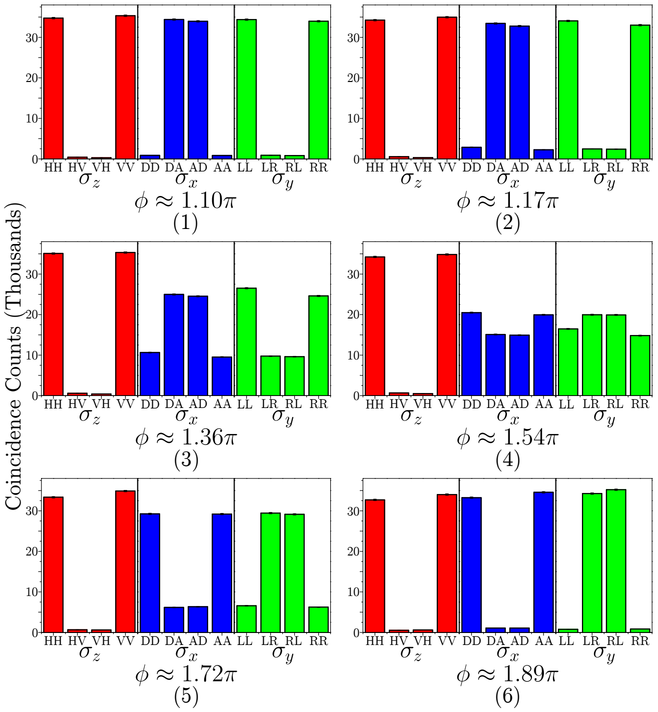


FIG. 4: (Color online.) Experimental setup for producing $|\Phi(\phi)\rangle = \frac{1}{\sqrt{2}}\left(|HH\rangle + e^{i\phi}|VV\rangle\right)$ polarization states. Photon pairs are generated via type-I noncollinear SPDC in a pair of orthogonally oriented BBO crystals and analyzed with wave plates and polarizing beamsplitters. The phase ϕ is adjusted by tilting a quarter-wave plate.



$\phi \approx 1.10\pi$
(1)

$\phi \approx 1.17\pi$
(2)

$\phi \approx 1.36\pi$
(3)

$\phi \approx 1.54\pi$
(4)

$\phi \approx 1.72\pi$
(5)

$\phi \approx 1.89\pi$
(6)

FIG. 5: (Color online.) Results of projective measurements on six states of the form $\frac{1}{\sqrt{2}}\left(|HH\rangle + e^{i\phi}|VV\rangle\right)$, corresponding to the eigenbases of the operators $\{\sigma_x \otimes \sigma_x, \sigma_y \otimes \sigma_y, \sigma_z \otimes \sigma_z\}$. The approximate value of the phase is included for each case. Counts were integrated over 1 s per measurement setting.

plate tilt range of twelve degrees and transforming the state from $|\Phi^-\rangle$ to $|\Phi^+\rangle$.

Projective measurements were taken in three bases, corresponding to the eigenbases of the operators $\{\sigma_x \otimes \sigma_x, \sigma_y \otimes \sigma_y, \sigma_z \otimes \sigma_z\}$. We will refer to the elements of these bases as $|x_i\rangle\langle x_i|$, $|y_i\rangle\langle y_i|$ and $|z_i\rangle\langle z_i|$ respectively. For example, the eigenbasis of $\sigma_z \otimes \sigma_z$ is given by $|z_1\rangle = |HH\rangle, |z_2\rangle = |HV\rangle, |z_3\rangle = |VH\rangle, |z_4\rangle = |VV\rangle$, and similarly for the other bases. To verify the entanglement of these states, an accessible nonlinear witness was constructed from the linear witness $W = (1/4)(\mathbb{1} + \sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z)$. Following [15], the expectation value $w_\infty(\sigma)$ of the nonlinear witness for a state σ can be expressed as

$$w_\infty(\sigma) = \mathrm{Tr}\left(\rho W\right) - |c|^2 - \frac{|d|^2}{1-|k|^2}, \tag{34}$$

where

$$\begin{aligned} c &= \mathrm{Tr}\left[\sigma(|\psi^-\rangle\langle\psi^-|U)^t\right] \\ k &= \mathrm{Tr}\left(\sigma U^t\right) \\ d &= \mathrm{Tr}\left(\sigma W\right) - ck, \end{aligned}$$

$|\psi^-\rangle = \frac{1}{\sqrt{2}}(|HV\rangle - |VH\rangle)$ and the superscript t denotes partial transposition. By choosing $U = \sigma_z \otimes \sigma_z$, this expectation value can be computed from the expectation value of the aforementioned operators and the nonlinear witness is *accessible* [15]. An accessible nonlinear witness was chosen because it detects these entangled states for most values of ϕ.

In this experiment, all measurements are independent so that each element of the POVM $\{B_i\}$ is a tensor product of the operators corresponding to possible individual outcomes. The likelihood function takes the form

$$\begin{aligned} \mathcal{L}_i(\sigma) = \prod_{j=1}^{4} &\mathrm{Tr}\left(\sigma|x_j\rangle\langle x_j|\right)^{n_x^j} \cdot \mathrm{Tr}\left(\sigma|y_j\rangle\langle y_j|\right)^{n_y^j} \\ &\cdot \mathrm{Tr}\left(\sigma|z_j\rangle\langle z_j|\right)^{n_z^j}, \end{aligned} \tag{35}$$

where n_x^j is the number of times outcome $|x_j\rangle\langle x_j|$ is obtained and similar definitions hold for the other operators, so that the total number of measurement outcomes is $n = \sum_{j=1}^{4} n_x^j + n_y^j + n_z^j$. Note that in this case the measurement outcome B_i is fully specified by the numbers $\{n_x^j, n_y^j, n_z^j\}$. In the experiment, six states were prepared corresponding to six different values of the parameter ϕ. The measurement outcomes for each case are summarized in Fig. 5.

We have calculated the confidence as in equation (16) for the six preparations of the entire experiment. These results are illustrated in Table I. We can report very high confidences for almost all states, with the exception of state 4 for which condition (8) cannot be satisfied for any value of ϵ. This is not entirely surprising as this state presents the weakest correlations in the $\{|x_j\rangle\langle x_j|\}$ and $\{|y_j\rangle\langle y_j|\}$ bases leading to a value of the nonlinear witness that is closest to zero, as seen in Table I. Thus, the outcomes for this case most closely resemble the ones that could be obtained from a separable state. This again is evidence that only large data which are clearly inconsistent with separable states can lead to the reliable statements obtained from our procedure.

Additionally, we are interested in understanding how the maximum achievable confidence depends on the total number of runs of an experiment. It is also important to

State	Approximate phase	Confidence	w_∞
1	1.10π	5150	-23.0
2	1.17π	2050	-15.2
3	1.36π	410	-3.4
4	1.54π	0	-0.3
5	1.72π	1819	-5.8
6	1.89π	4980	-13.6

TABLE I: Calculation of the confidence and value of the nonlinear witness for all prepared states in the experiment. The total number of counts obtained in each case was roughly 35,000.

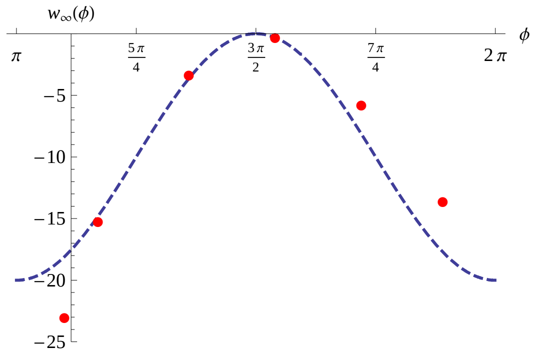


FIG. 6: (Color online) Value of the nonlinear witness $w_\infty(\phi)$ for the six states prepared in the experiment (dots). The value of the nonlinear witness for the family of states $\sigma(\phi) = (1-p)|\Phi(\phi)\rangle\langle\Phi(\phi)| + \frac{p}{4}\mathbb{1}$ with $p = 1/42$ is shown in the background (dashed). This curve is included only to illustrate the values of ϕ for which it is difficult to verify entanglement and should not be interpreted as a fit to the data. The value of p was chosen to adjust the scaling to the recorded values.

Total counts	Confidence (Γ_α)	Confidence (Γ_W)
1500	0	0
3000	18	24
6000	165	200
15000	300	315
30000	660	700
60000	1378	1500

TABLE II: Calculation of the confidence for samples of different size from the outcomes of experiment 6 based on Γ_W and Γ_α.

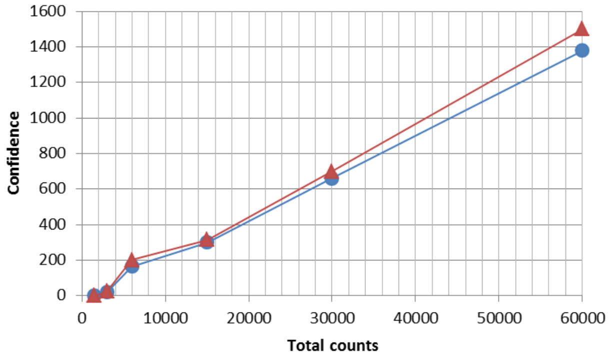


FIG. 7: (Color online) Confidence for random samples of different size, quantified by the total number of counts. The confidences were calculated for Γ_W (triangles) and Γ_α (dots). These results illustrate that the bound introduced by considering the subset Γ_α is small and is not an impediment to reach a very large confidence. In the case of 1500 total counts, the confidence is zero, consistent with our understanding that a large number of outcomes are needed in order to reliably report entanglement with our technique. Moreover, the data shows that the confidence is roughly linear in the number of outcomes.

gain insight on the cost of using the bound of Observation 1 for linear witnesses. For this purpose, samples of different size were randomly selected from the outcomes of experiment (6) in Fig. 5. That is, from the entire set of observations in this experiment (shown in Fig. 5), we randomly selected a subset of all the data and interpreted it as arising from an experiment with a fewer number of runs (counts). The confidence was calculated for each of them using both regions Γ_W and Γ_α, this latter being possible because this state is also detected by the linear witness. The obtained values using these two different methods is portrayed in Fig. 7 and Table II.

The results indicate that, as a percentage of the total confidence, the loss introduced by considering Γ_α is small. It is also clear that a large number of runs are necessary in order to report a non-zero confidence, in accordance to our understanding of the role of the enlarging parameter δ and the polynomial factor $c_{n,d}$ as discussed in section II. To estimate the numerical error present in the SA algorithm, we performed 20 independent runs of the algorithm for the data of state 1 and found this numerical error to be 1.85%. In all calculations it was ensured that condition (8) was satisfied by at least ten times this numerical error. In the construction of the integration regions a value of $\eta = 10^5$ was chosen for all cases. Finally, the CVX package for specifying and solving convex programs [35] was used to numerically calculate the global maximum of the likelihood function, as well as its maximum over Γ_α.

VI. CONCLUSION

In this paper, we have applied the work of M. Christandl and R. Renner in Ref. [1] to the case of entanglement verification. Through the concept of confidence regions, we have provided a procedure to make reliable and efficient statistical statements quantifying the confidence level of having entanglement present in a physical system. These statements have a clear operational interpretation and do not require the specification of a prior distribution nor the assumption of independent measurements or i.i.d. sources. We have shown that this method can be applied in practice by developing specific numerical tools designed to calculate all necessary quantities. For the particular case of experiments relying on linear entanglement witnesses, we have shown that the procedure can be implemented efficiently using only plain Monte Carlo integration and convex optimization methods. The procedure is ready to be applied to current experiments as we demonstrated by applying the technique to data obtained from an experiment generating entangled two-photon states. High confidence values can be achieved whenever the data is strongly inconsistent with a separable state and the number of measurement outcomes is large enough. Our results thus provide an illustration of the techniques that must be employed in current experiments in order to obtain clear and reliable claims.

It is important to note that this work assumes that there are no systematic errors in the measurements performed. In any real experiment, there will always be discrepancy between the intended measurement and the one actually performed, no matter how small this discrepancy is. These systematic errors can in principle lead to incorrect statements and a method to incorporate it in the framework must be pursued. Numerical techniques also invariably involve errors and these should also be clearly incorporated in the framework. Future research may lead to improved algorithms. Finally, let us note that it is often desirable to quantify the amount of entanglement present as opposed to just verifying it. Our technique can in principle be applied to such cases by reporting regions that contain states with at least a certain amount of entanglement. Future work can focus on including this feature into the procedure.

Acknowledgements.- We thank Chris Ferrie for his insight on numerical methods to integrate probability distributions and Matthias Christandl and Renato Renner for their hospitality in Zürich and fruitful discussions concerning their work. Most importantly, we thank Philippe Faist for valuable discussions on his work and on convex optimization techniques. Oleg Gittsovich is grateful for the support of the Austrian Science Fund (FWF) and Marie Curie Actions (Erwin Schrödinger Stipendium J3312-N27). This work was supported by Industry Canada, Ontario Ministry of Research and Innovation ERA, QuantumWorks, Ontario Centres of Excellence, the Canadian Foundation for Innovation, and NSERC Discovery and NSERC Strategic Project Grant (SPG) FREQUENCY.

Appendix

Here we fully describe the simulated annealing (SA) algorithm. The algorithm is based on a biased random walk in state space that preferentially selects states with a higher value of the likelihood function at each new step of the iteration. However, it also accepts jumps to states with lower values with a probability that depends on a global parameter T, usually referred to as the temperature because of its similarity with the physical temperature in the annealing process of metallurgy.

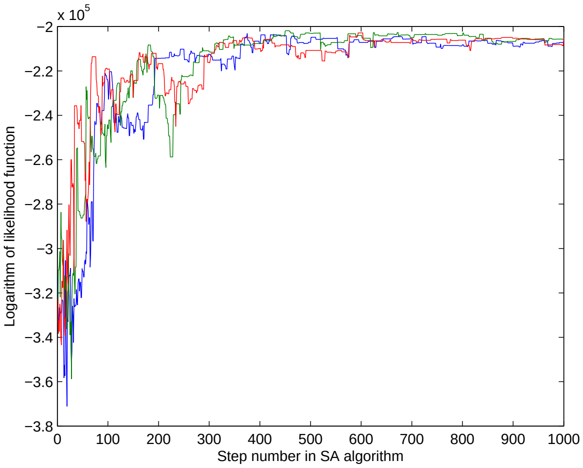


FIG. 8: (Color online) Three independent runs of the same simulated annealing algorithm for the data of experiment 1. Although all parameters are identical in each case, the output is slightly different in each case due to the stochastic nature of the algorithm.

Below is a full enumeration of all the steps of the algorithm to calculate the maximum value of the likelihood function $\mathcal{L}(\sigma)$ over the set Γ_W. A graphical illustration of how the maximum value of the function is reached as the algorithm progresses is found in Fig. 8. The random walk here described is based upon the quantum adaptation of the Metropolis-Hastings algorithm depicted in [32].

Simulated annealing algorithm:

1. Select an initial value T_0 for the temperature T as well as for the "step size" Δ.

2. Generate a $d \times d-$dimensional random state $|\psi\rangle$ according to the Haar measure, where d is the dimension of the underlying Hilbert space $\mathcal{H}$. Trace out one of the subsystems to obtain the state σ_0. If $\sigma_0 \in \overline{\Gamma_W}$ continue to the next step, repeat otherwise.

3. Randomly choose a 2×2 Hermitian matrix H_{kl} in the following way. Pick two integers k, l randomly from the set $\{1, 2, \ldots, d\}$. If $k < l \to H_{kl} = |k\rangle\langle l| + |l\rangle\langle k|$, similarly if $k > l \to H_{kl} = -i|k\rangle\langle l| + i|l\rangle\langle k|$ and finally if $k = l \to H_{kl} = |k\rangle\langle k| - |k+1\rangle\langle k+1|$ (set $k + 1 = 1$ if $k = d$).

4. Pick a distance δ by sampling from a Gaussian distribution with mean 0 and standard deviation Δ.

5. Compute the state $|\psi'\rangle = \exp(iH_{kl}\delta)|\psi\rangle$. Trace out one of the subsystems of $|\psi'\rangle$ to obtain the state σ_0'.
6. If $\sigma_0' \notin \overline{\Gamma_W}$, repeat steps 2 to 5, continue otherwise.
7. Evaluate the ratio $R = \log\left(\mathcal{L}(\sigma_0')/\mathcal{L}(\sigma_0)\right)$. If $R > 0$ ($\mathcal{L}(\sigma_0') > \mathcal{L}(\sigma_0)$), let $\sigma_1 = \sigma_0'$. Otherwise, flip a coin with bias $p = \exp\{-|\log(\mathcal{L}(\sigma_0)) - \log(\mathcal{L}(\sigma_0'))|/T\}$. If "1" is obtained (which happens with probability p), again let $\sigma_1 = \sigma_0'$, otherwise $\sigma_1 = \sigma_0$.
8. Repeat steps 2-6 N times to generate a set $\{\sigma_1, \sigma_2, \ldots, \sigma_N\}$ corresponding to N steps of the random walk. For each step, adapt the temperature via the cooling rule $T(s) = T_0/s$ where s is the step of the walk. The maximum value of $\mathcal{L}(\sigma)$ over this set is the output of the algorithm.

The performance of the algorithm depends strongly on the value of Δ and this value must be adapted throughout each step of the walk in order to maintain a fixed average acceptance ratio, i.e. the fraction of times we jump to a new state. Various values for these ratios are suggested [36]. Similarly, the choice of initial temperature is crucial. Its role is to prevent the algorithm from being stuck in local maxima by allowing it to escape such cases in the initial stages of the algorithm. The temperature is then reduced to ensure that convergence to the maximum is attained. Therefore, the choice of initial temperature and cooling rule is essential and varies for different cases. In practice, they must be chosen for each particular problem based mostly on experience.

Finally, in order to check whether a new state belongs in $\overline{\Gamma_W}$, it is necessary to determine the maximum fidelity of this state with any state in this set. For this purpose, we exploit the fact that the fidelity function is concave in both its arguments and that the restriction $\rho \in \overline{\Gamma_W^\delta}$ is convex for both linear and nonlinear witnesses. These properties allow us to employ the highly efficient tools of convex optimization to solve the maximization problem. Concretely, for a given state σ, we verify membership in $\overline{\Gamma_W}$ by solving the problem

$$\begin{aligned} \text{maximize} \;\; & F(\sigma, \sigma') \\ \text{subject to} \;\; & \sigma' \in \overline{\Gamma_W^\delta} \end{aligned}$$

where σ' must be forced to be a density operator. The state σ is a member of $\overline{\Gamma_W}$ if the solution to this problem is larger than $\sqrt{1-\delta^2}$. In our case, the CVX package for specifying and solving convex programs [35] was used to numerically solve the problem.

[1] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012).

[2] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).

[3] M. Curty, M. Lewenstein, and N. Lütkenhaus, Phys. Rev. Lett. **92**, 217903 (2004).

[4] G. Vidal, Phys. Rev. Lett. **91**, 147902 (2003).

[5] O. Gühne and G. Tóth, Physics Reports **474**, 1 (2009).

[6] N. Killoran, M. Hosseini, B. C. Buchler, P. K. Lam, and N. Lütkenhaus, Phys. Rev. A **86**, 022331 (2012).

[7] W. Gao, C. Lu, X. Yao, P. Xu, O. Gühne, A. Goebel, Y. Chen, C. Peng, Z. Chen, and J. Pan, Nature Physics **6**, 331 (2010).

[8] T. Moroder, M. Kleinmann, P. Schindler, T. Monz, O. Gühne, and R. Blatt, arXiv preprint arXiv:1204.3644 (2012).

[9] S. J. van Enk, N. Lütkenhaus, and H. J. Kimble, Phys. Rev. A **75**, 052318 (2007).

[10] R. Blume-Kohout (2012), e-print arXiv:1202.5270.

[11] R. Blume-Kohout, New J. Phys. **12**, 043034 (2010).

[12] B. Jungnitsch, S. Niekamp, M. Kleinmann, O. Gühne, H. Lu, W. Gao, Y. Chen, Z. Chen, and J. Pan, Phys. Rev. Lett. **104**, 210401 (2010).

[13] W. Eadie, D. Drijard, F. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North-Holland Publishing Co., 1971).

[14] R. Blume-Kohout, J. O. S. Yin, and S. J. van Enk, Phys. Rev. Lett. **105**, 170501 (2010).

[15] J. M. Arrazola, O. Gittsovich, and N. Lütkenhaus, Phys. Rev. A **85**, 062327 (2012).

[16] G. Chiribella, Theory of Quantum Computation, Communication, and Cryptography Lecture Notes in Computer Science **6519**, 9 (2011).

[17] K. Zyczkowski and H.-J. Sommers, J. Phys. A: Math. Gen. **34**, 7111 (2001).

[18] M. Christandl, Private communication (2012).

[19] L. Gurvits, in *Proceedings of the thirty-fifth ACM Symposium on thory of computing, San Diego, CA* (San Diego, CA, USA, 2003), p. 10.

[20] D. Bruß, J. I. Cirac, P. Horodecki, F. Hulpke, B. Kraus, M. Lewenstein, and A. Sanpera, J. Mod. Opt. **49**, 1399 (2002).

[21] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).

[22] B. Terhal, Phys. Lett. A **271**, 319 (2000).

[23] M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki, Phys. Rev. A **62**, 052310 (2000).

[24] M. Bourennane, M. Eibl, C. Kurtsiefer, S. Gaertner, H. Weinfurter, O. Gühne, P. Hyllus, D. Bruss, M. Lewenstein, and A. Sanpera, Phys. Rev. Lett. **92**, 087902 (2004).

[25] O. Gühne and N. Lütkenhaus, Phys. Rev. Lett. **96**, 170502 (2006).

[26] T. Moroder, O. Gühne, and N. Lütkenhaus, Phys. Rev. A **78**, 032326 (2008).

[27] M. Agnew, J. Z. Salvail, J. Leach, and R. W. Boyd, *Entanglement verification via nonlinear witnesses*, arXiv:1210.1054 (2012).

[28] C. Fuchs and J. Van De Graaf, Information Theory, IEEE Transactions on **45**, 1216 (1999).

[29] Bengtsson and Życzkowski, *Geometry of Quantum States* (Cambridge University Press, 2006).

[30] Z. I. B. Dirk P. Kroese, Thomas Taimre, *Handbook of Monte Carlo Methods* (John Wiley & Sons, 2011).

[31] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein, Phys. Rev. A **58**, 883 (1998).

[32] R. Blume-Kohout, New J. Physics **12**, 043034 (2010).

[33] P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P. H. Eberhard, Phys. Rev. A **60**, R773 (1999).

[34] J. Lavoie, R. Kaltenbaek, M. Piani, and K. J. Resch, Phys. Rev. Lett. **105**, 130501 (2010).

[35] I. CVX Research, *CVX: Matlab software for disciplined convex programming, version 2.0*, `http://cvxr.com/cvx` (2012).

[36] S. Chib and E. Greenberg, The American Statistician **49**, 327 (1995).